DOCUMENT RESUME

ED 324 956 FL 018 941

AUTHOR Fjeld, Ruth Vatvedt
TITLE Readability of Legislative Texts and Public Information.
PUB DATE Apr 90
NOTE 14p.; Paper presented at the Meeting of the World Congress of Applied Linguistics sponsored by the International Association of Applied Linguistics (9th, Thessaloniki, Greece, April 15-21, 1990).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Discourse Analysis; Foreign Countries; *Hearings; *Information Dissemination; Language Research; *Legislation; *Norwegian; Policy Formation; Public Administration; *Readability; Reading Research; Taxes; Technical Writing; Uncommonly Taught Languages
IDENTIFIERS *Norway

ABSTRACT

A study investigated the readability of Norwegian law texts intended for both the legal profession and the public (e.g., laws regulating social insurance and public administration) that contained public information about tax payment. Six passages from the samples were rewritten by changing a number of specific morphological, lexical, and syntactic items. Four samples were rewritten in three versions, with changed lexical items, changed syntactic items, and changed lexical and syntactic items). The texts were presented to 28 well-educated, non-expert readers employed in government administration. Reading time for each version, controlled for individual reading speed, was measured. All versions were read by at least six readers. The readers were then asked content, comprehension, and structural questions about the passages. In the case of two texts, readers were asked for their opinions of the readability. Results indicate that for all of the texts, answers to content questions were best on the versions in which both lexical and syntactic items were changed. Results for other adapted versions and for reading time are less clear. Results of a computer analysis of the original texts' item frequency and distribution suggests a mixed writing style that probably does not contribute to readability. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED324956

READABILITY OF LEGISLATIVE TEXTS AND PUBLIC INFORMATION

Ruth Vatvedt Fjeld

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Fjeld

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL018941

April 1990

Ruth Vatvedt Fjeld,
University of Oslo,
Department of Scandinavian Studies and Comparative Literature,
Section for Norwegian Lexicography:

READABILITY OF LEGISLATIVE TEXTS AND PUBLIC INFORMATION

Readability can be investigated from many points of wiew. For me it is important to find a way to study the reading problems through linguistic features, and to leave social or psychological features to other professionals. But the main problem then is to know whether it is linguistic features that cause an eventual improvement of the readability of a text. And if that is the case, what are those linguistic features? Language is a complicated matter, and so is the reading process, and you never feel sure you can prove anything.

The background for my investigation is an earlier work of mine, dealing with functional illiteracy among adults in Norway. This work showed a clear correlation between functional illiteracy and complicated or badly written texts. My study presupposes that it is possible to present the same content in different versions, some presumably easier to read than others.

It is practical to arrange the linguistic features that are to be studied on different levels. The traditional ones are

single words
phrases
sentences

Some studies also include the text level, or discorse studies. Besides these levels, we have to concern ourselves with the level of interpretation, that is, how the chosen linguistic factors make the reader understand or not understand the text.

The first problem to deal with, is to isolate the features that are to be studied. It is not enough to investigate what a reader understands from an isolated word, you have to put the words into a context. But if you do, how can you know that the problems for the reader are attached to the actual word and not to something else in the text?

The most common way to solve this problem is to construct different versions of a text, where only the actual word has been changed. Still there is some uncertainty. Maybe the change in the result has been caused by the new word you put into the text, and not by the old one you took out?

To help you out of such problems, you must trust your own linguistic intuition. To some degree you can get support from frequency investigation and etymological clues.

My investigation examines some law texts meant for both the

legal profession and the public, for example laws regulating social incurance, and laws regulating public administration rules. In addition to the laws, I have investigated some public information about tax-payment, which has to be understood by every working person in the country.

Some of the texts are law texts written for professionals, and one text I have taken from a parliamentary bill also written for professional administrators.

The method I have used is to rewrite some text passages by changing the linguistic entities that I want to study, to see if these have some relevance to readability. The entities are categorized into

morphological items:
- morphological variants of radical or moderate form of Norwegian bokmål
- nominal phrases with or without article

lexical items:
- vague forms and expressions
- loan words and other foreign words
- professional words or official jargon
- archaic words

syntactical items:
- passive voice
- word order
- nominalization
- sentence complexity

Morphological items:

Common forms

A very special aspect of the Norwegian language is its two standards of modern Norwegian, bokmål (book-language) and nynorsk (Neo-Norwegian). Bokmål is spoken by most people in Eastern Norway, nynorsk in the west. Also bokmål is much influenced by Danish, a reminiscence of our history as a Danish province. Bokmål is therefore sometimes called "provincial Danish". It should also be added that bokmål is the prestige standard of Norwegian today.

Some language planners in Norway have for several decades tried to make one common standard out of the two, a very obvious solution to many linguistic and social problems that the two standards represent to a small society, since those two standards are linguistically very closely related. The common form is called samnorsk (common Norwegian), but it is yet no fully elaborated standard, in fact only a list of word-formations. However, these samnorsk-forms (the common forms) have mostly been rejected by the language users. The writers write either pure nynorsk, or pure bokmål. Some few, linguistically very conscious writers all the same succeed in writing the samnorsk variety, but then with a syntax close to the Neo-Norwegian, and a consequent choice of radical word forms. The writers in the official bureaucracy had to follow,

or often found it convenient to follow the advice to use the common forms, but without changing anything else in their writing style. Hence the common forms today are more or less a stylistic feature of public information. On many occasions the common forms have created a new terminology for new legislative or bureaucratic phenomenons, such as

attføring - rehabilitation (rehabilitering)
bevilling - licence (bevilgning)
tilskott - contribution (tilskudd)
stønad - benefit, aid (støtte, understøttelse)

In the more traditional bokmål these words would have been the forms listed on the right. As most writers of public information, especially the law makers, usually write a conservative style, as to lexical forms as well as syntactic constructs, these common forms represent an abrupt and striking change in style. Psycholinguistically these forms are a provocation to many conservative readers, as the rest of the style is to the more radical ones. This mixed style also makes a bad style, it is difficult to imagine a normal, living person writing in this way. My hypothesis is therefore that frequent use of common forms combined with conservative linguistic style is influencing the readability of the text.

Another way of chosing common forms in Norwegian is by the morphemes -ing, as a radical form, for nominalizations, or -else, which is a conservative form. An automatic counting of what the authors choose, can be done by text analyzing programs. I have used the TACT-program developed by John Bradly, University of Toronto, analyzing some entire law texts. The results show that -ing is a much more frequent morpheme for nominalizing than -else, but it also shows that many of the -else forms can be changed into -ing forms, which the authors also sometimes do. For example is a form like fastsettelse (stipulation) used 16 times in the law of social security, while fastsetting is used 4 times in the same text. The same goes for feriegodtgjørelse/-ing, overtredelse/-ing, utdannelse/-ing. For most of the forms with -ing, this is the only possible form in modern Norwegian, while a very few forms with -else are bound to this morpheme - in most of these cases the authors are free to choose, which means that this is a matter of style, and the lawmakers here often choose the most conservative forms possible. It seems to me, that the common forms are preferred only when they have ben lexicalized and have got a special meaning or stylistic value, otherwise the conservative forms are chosen.

Articles

Another typical feature for Norwegian is the double article in defining the definite form of nouns:

de samme reglene (the same rules+def.article)

while Danish and conservative or archaic Norwegian have

de samme regler (the same rules).

Besides this, legislative texts often use nominals with no article at all, so-called naked nouns

> Dersom person som går inn under reglene for ...
> If person comes under the rules of ...

However, structural features like articles give the reader some clues as to how the constituents are formed. This has been tested and proved in English by among others Epstein 1961. M. Pinkal (1985) also points at the fact that definite articles have an indexing function, they function as markers or quantifiers to make the nominals less vague.

All types of article-reducing imply a distance to natural, modern style of prose. I therefore put forth the hypothesis that texts with few or no articles in the NPs are less readable than texts with more articles, both the double Norwegian article, and the normal article.

Lexical items:

Vague forms

Vagueness is a difficult matter to investigate. Few redability investigations have dealt with the problem of vagueness. Maybe it is because it is so difficult to classify. One of the best and most useful classifications I have found, was presented by Kempson in her Semantic Theory of 1977. She lists four types of vagueness

> (i) referential vagueness, where it is unclear what an item refers to in the real world, that is, if an item can be used for an object or not (example: city)
>
> (ii) indeterminacy of meaning - where the meaning of the word might be changed according to context and situation (example: good)
>
> (iii) lack of specification - where the meaning is clear but very general (example: go, do, neighbour)
>
> (iv) the meaning of an item involves the disjunction of different interpretation

Hiller & al (1968) made a study of vagueness in an investigation of readability. They classified vagueness as indeterminate qualifiers (rather, very, any) or probability (could be, might), and found that the proportion of these words had a negative correlation with the difficulty of the texts studied.

In law texts vagueness is often expressed in normal adjectives as

longlasting sickness
appropriate treatment
important medicines

or nouns/noun phrases
in connection to
appropriateness
relationship

or verbs

to be included into
to go in under
to be regarded as

or modal verbs

might make exceptions
the department might make rules

All these vague expressions leave a subjective desicion to the reader or user of the text. Whether the rule is to be used or not, is a matter of judgement for the reader. I have found it helpful to use the categorization presented by Kempson, because she also includes lack of specificity in her system. I have found no better way to register vagueness than to examine the texts closely and mark each occurrence of vague items with a special code. Then it is possible to count the frequency of the occurrences and classify the different types of vagueness. Later on I hope to find a way to systematize these findings and make a basis for automatic registration.

Loan words

Loan words are defined as words that etymologically are non-Norwegian and that can still be easily recognized as such.

Professional terms

Professional terms are mostly loan words that do not belong to the general knowledge of non-professionals, but also heritage words used in a special professional meaning are counted here.

Archaic forms

It is well known that officialese often prefer archaic forms, many words live their own lives in official documents. Some of the words are purely lexical words, easily replaced by more modern ones, as nedkomst (delivery) instead of fødsel (birth), or tarv (demands) instead of behov (needs). But many forms are more deeply integrated in the text, which means that they are structure words, such as pronominal adverbs as herved (hereby, herewith) herpå (hereupon). To replace these words with more modern ones, often means to replace entire syntactic constructs, so that the rewriting leads to profound changes of the texts.

I also include official jargon in the cathegory of archaic forms. Frequent use of these forms in combination with frequent use of common forms, is a typical marker of the mixed style I am interested in.

Syntactic items

Passive voice has in many investigations and from different points of wiew been regarded as less easily understood than the active voice. I therefore include this factor in my investigation.

Word order is complicated in legal writing, and often deviates from the normal word order because of perspective marking or special focusing. Often the word order is unnatural because of many interpolated reservations, conditions and the like. It is impossible to rewrite a text without making changes in the word order.

Nominalization is a standard feature in most readability testing, and has therefore been included here

Sentence complexity makes it possible to include many of the other syntactic features that are regarded as markers of a difficult text. This shows why sentence length is important to readability. It is not the length alone that makes the text difficult, the length is the consequence of the other features.

I have not tried to do anything original or new in my syntactical study. The purpose of this part is to show the relation between syntactic, lexical and morphological features, and how they influence the readability of the text.

TEST METHOD

I have chosen six short texts for rewriting. Four of them have been rewritten in three versions

version 0 - the original
version I - changed lexical items
version II - changed syntactical items
version III - both lexical and syntactical items have been changed

As far as possible I have tried to find texts where the interesting features are fairly well spread throughout the text. The features have been counted and classified as shown in table 1.
Two of the texts have been rewritten in order to investigate the common forms and the articles, and they therefore occur in only two versions. In this way I hope to have isolated the linguistic features that I want to study. This represents a different method than the one used with the other 4 texts, where the influence of the factors have been accumulated in

the same version.

The texts have been presented to 28 non-expert readers, well educated persons employed in governmental administration (8 employees from the Directorate of Labour, 24 from the Directorate for Civil Defence and Emergency Training, 1 physician), 15 women and 18 men, their age varying from 23 to 62 years.

Reading time are measured for the different versions. To neutralize the subjects different reading speed, I found the average reading speed for each of them, and calculated the single results for the different texts in percent of this average. In that way it is possible to compare the results in reading time from person to person. The main problem in all readability testing is that you cannot give the different versions of one text to one and the same person to see if he reads the one better than the other, or faster than the other. I have systematically arranged the different versions so that all versions have been read by at least 6 subjects.

After the subjects had read a text, they had to answer some control questions about the content, some multiple choice questions, some open questions and in two cases they had to reconstruct the content of the passages they have read.

RESULTS

Table 2.

The test showed that for all the texts the results for the control questions were best for the version where both syntax and lexical items had been changed. Where only syntax or only lexical forms had been changed, the results are not so clear. Three of the texts give better results with changed lexical items, two no change or nearly none, one gives a worse result. For the syntactical changes, the results are marginally better.

As to the results for reading time, they also are more or less dubious.

For the two texts measuring the influence of the articles or the use of common forms I also asked the subjects to give their opinion of the readability of the texts. These evaluations show a better result for the rewritten versions.

I therefore conclude that rewriting texts in respect of vague expressions, professional terminology and loan words show a tendency of bettering the readability. The same goes for syntactic features like nominalization, odd word order, passive voice, sentence complexity, when the lexical items are also changed. To obtain more secure evidence, one would need a more comprehensive investigation.

Of greater interest are the findings showing that to add articles to a law text can make people find it easy to read, while just the same text without these articles are estimated

as difficult. The same tendency can be extracted for texts with or without frequent use of common forms. This correlates with the findings of the other texts, where both the single words and the syntax had to be changed in order to show some improvement in readability. The changes in text E are also a change from a mixed style to a more uniform one.

Table 3.

Relative reading rate shows a more untidy picture. The reason is partly that my sample is to small, and also that there arose problems in the timing so that many results were lost. Anyway I find reading rate a very dubious as a measure of readability, and it requires a very expensive testing system with many tested persons to compensate for individual reading speed and differences between the texts.

THE TEXT ANALYZING PROGRAMME

The test is first of all meant to give evidence for the relevance of readability for the items that I examine by coding natural texts for frequency analyzes. As mentioned, I have decided to use the TACT programme for this purpose.

I first go into a text and give codes to the items I want to registrate. This coding is an adapted version of the SGML-method (Standard Generalized Markup Language) developed by the international project The Text Encoding Initiative. In this way it is possible to get KWIC concordances and frequency countings and also tables showing how the items are distributed througout the texts. It is also easy to see changes in style, in this work it means changes between radical forms and conservative ones. It is striking to see how the different forms occur in the same text, showing the insecurity of the authors. For instance one of the law text shows this variation :

common form	freq	conservative form	freq
heimel	5	hjemmel	7
sein	5	sen	4
arbeidsløyse	6	arbeidsløshet	0
		skjødesløshet	1
arbeidslaus	0	arbeidsløs	12
sjukdom	0	sykdom	56
stønad	87	støtte	2
framtidig	19	fremtidig	4
hensyn	19	omsyn	4
for-/tilskott	25	for-/tilskudd	9
nytte	13	benytte	3

All these forms are allowed within the standard of bokmål, and it is difficult to see why some radical forms have been more used than others. It might be caused by linguistic factors, as I presume that the form sjukdom is not used because of its phonetic form, but more probably forms which occur in a law written in Neo-Norwegian more easily spread to the bokmål.

The result, anyway, is a very special mixed style that gives the reader the inevitable feeling of reading a piece of officialese, and I do not think that this style makes official information easier to comprehend, not to say follow.

TABLE 1

Number of lexical changes in per cent of text length

Text	Number of words in version 0	added articles	common forms	cliches	archaic forms	loan words	professional terms	vague expressions	other changes	total number of changes	rank
A	78	0	2,56	2,56	11,55	0	1,28	2,56	2,56	23,07	2
B	79	0	1,27	1,27	6,33	0	0	0	2,54	11,41	4
C	99	0	0	2,02	14,14	1,01	7,07	6,06	0	30,30	1
D	169	0	0,59	4,73	5,33	4,14	0	2,96	0,59	18,34	3
E	118	-	9,7								
F	133	6,8									

Number of syntactical changes in per cent of text length

Text	passive → active	word order	nominali-zations	sentence complexity	others	total	rank
A	5,13	2,56	1,28	0	0	9,0	1
B	1,27	2,53	1,27	3,80	0	8,87	2
C	1,01	0	1,01	2,02	3,03	7,07	3
D	0	1,18	1,18	1,78	1,18	5,32	4

TEXT	No. control questions	PERCENTAGE OF CORRECT ANSWERS TO CONTROL QUESTIONS			
		Version 0	Version I	Version II	Version III
A	8	28% + 14 (n=9)	27% + 8 (n=8)	30% + 15 (n=8)	31% + 12 (n=8)
B	6	46% + 21 (n=8)	70% + 32 (n=9)	65% + 29 (n=8)	73% + 28 (n=8)
C	3	58% + 46 (n=8)	58% + 30 (n=8)	62% + 45 (n=7)	63% + 54 (n=9)
D	5	63% + 24 (n=7)	51% + 20 (n=9)	63% + 23 (n=8)	67% + 25 (n=9)
E	3	61% + 33 (n=6)	86% + 30 (n=12)		
F	3	83% + 27 (n=12)	97% + 10 (n=11)		

Readability expressed as the fraction of correct answers to control questions about the text. Results from 6 different texts (A - F) written in different versions are presented. Mean and standard deviation. Number of persons in parenthesis.

TABLE 3

Relative reading rate in per cent of mean reading rate for each subject

Text	version 0	version I	version II	version III
A	95 (n=5)	106 (n=6)	120 (n=2)	122 (n=2)
B	96 (n=4)	73 (n=6)	79 (n=7)	111 (n=3)
C	77 (n=4)	82 (n=5)	87 (n=4)	83 (n=6)
D	122 (n=4)	117 (n=5)	122 (n=4)	114 (n=4)
E	89 (n=3)	83 (n=5)		
F	131 (n=5)	116 (n=7)		

END

U.S. Dept. of Education

Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 29, 1991